Smart Systems, Security, and Data Science (SSSDS) **2025, 1(1): 1-10**
https://www.longmanpublishers.com/

An Ensemble Stacking Algorithm to Increase Bankruptcy Prediction Model Accuracy

Iskandar
International Islamic University Malaysia (IIUM), Kuala Lumpur, Selangor 53100, Malaysia

Abstract: To predict bankruptcy, bankruptcy analysis is necessary. Predicting bankruptcy incorrectly frequently results in bankruptcy. High-accuracy machine learning for reversal analysis has to keep getting better. To forecast bankruptcy, numerous machine learning algorithms have been used. To increase forecast accuracy, model improvisation is still required. We suggest a combination model based on the stacking ensemble approach and genetic algorithm-support vector machine (GA-SVM) to increase the accuracy of bankruptcy prediction. The Taiwan Economic Journal's Taiwanese Bankruptcy dataset is used in this study. Next, in order to deal with unbalanced datasets, we employ a synthetic minority over-sampling strategy. We use GA-SVM to choose the best feature, stack the classifier as a novel approach, and employ extreme gradient boosting as a meta-learner. With an accuracy of 99.58%, the results demonstrate the better precision achieved by the GA-SVM based on the stacking model. Compared to using a single classifier, the accuracy achieved is higher. This study demonstrates that the suggested approach has a higher accuracy rate in predicting bankruptcy.
Keywords: Bankruptcy prediction, Taiwanese Bankruptcy, Genetic algorithm, Stacking ensemble, SMOTE.

1. Introduction

Numerous crises affecting financial failure or bankruptcy have the potential to completely destroy the world economy (Abdillah, 2020). It may have a negative impact on corporations, small company owners, traditional marketplaces, and the government. Researchers, policymakers, investors, and entrepreneurs are trying to identify the causes of bankruptcy (Liang et al., 2016). A few things that can cause bankruptcy are the price of raw materials, employee compensation, the competitiveness of the business, and managerial ineptitude. Any business owner could face bankruptcy for a variety of reasons. Every level of entrepreneurs may be impacted by it (Blazy & Stef, 2020). Credit risk assessment can be improved by managing finances and reducing economic credit risk (Umar et al., 2021). The information offered by bankruptcy assessments may help governments, shareholders, management, and investors protect their finances from bankruptcy. Through bankruptcy study, early warning indicators and financial weak points can be found. Nonetheless, there are certain advantages to studying bankruptcy, such as reduced expenses for credit investigation, financial tracking, and collection rates.

Information on the company's cash flow and net income may be found in the financial statements. The corporation may be dissolved at the local government's discretion if cash flow is less than the target than the cost of capital provided and unpaid debt results in liquidity. If the business solely depends on state subsidies, even the state economy's collapse could hasten its demise (Foerster et al., 2017).

The economy, shifts in commodity prices, the country's surplus and deficit, and even the strength of the local currency can all be taken into account when determining the reasons behind bankruptcy. Socially, the company's target market will be impacted by shifts in local customs. Liquidation will result if the company is unable to adapt and change. If the company doesn't upgrade its systems, it will lose its ability to employ media and technology to its full potential. This may have an impact on the company's equipment upkeep; if not, an integrated system will just make things more challenging for the business. Removing tariffs and subsidies on the flow of import and export sales is the goal of governmental regulations (Hickel et al., 2022). Insolvency can also be caused by poor business management (Kücher et al., 2020).

To enhance financial management, financial bankruptcy analysis is required (Muslim & Dasril, 2021). To study financial bankruptcy, metrics and indicators are needed (Kozlovskyi et al., 2019). Six groups of financial ratios (FRs) were developed by Bateni and Asghari (2020) as indicators of bankruptcy prediction after they started a bankruptcy analysis of 158 companies. These groups include cash flow to total debt, net income to total assets, accumulation of current and long-term liabilities to total assets, current ratio, capital employment to total assets, and the period without credit. Almamy et al. (2016) conducted additional study on financial bankruptcy and discovered that five categories of financial ratios—cash flow to total debt, net income to total assets, total debt to total assets, working capital, and the current ratio—are utilized as predictors of bankruptcy. Leverage and liquidity are two other ratios that may bolster the necessity for study (Jumanto et al., 2023). The z-score is then used to refer to the bankruptcy prediction indicator. According to research by Liang et al. (2016), FRs and

Corporate Governance Indicators (CGIs) are two important considerations when examining insolvency. FRs factors include growth, turnover, cash flow ratio, capital structure ratio, profitability, and solvency. Board structure, ownership structure, cash flow rights, and retained specialists are all CGI considerations. Despite the large dimensions obtained, the model's performance will be developed by these two criteria. Indicators of financial insolvency will develop a pattern, and only analyzing human assumptions is insufficient (Muslim & Dasril, 2021). Models to reduce bankruptcy include machine learning and statistical analysis (Lin et al., 2019). Combining FRs allows for discrimination against financial distress, and different discriminant analyses can be used to identify particular ratios. It is insufficiently strong, though, to demonstrate that MDA alone can yield the optimal outcomes (Liang et al., 2020). Similar to a categorization model, bankruptcy analysis maps traits and indications of the reasons for bankruptcy using statistical data supplied by the business. Feature selection is applied based on the kind of data object that is owned. Labeled data is referred to as supervised data. One variant of the supervised feature selection technique that can lower high dimensions is the wrapper method (El Aboudi & Benhlima, 2016). By using the principles of crossover, mutation, and selection, the genetic algorithm (GA) exhibits natural dimensions. GA characteristic rules can be used to create discriminant classifiers and solve nonlinear optimization issues. In the feature selection process, GA can be applied as a wrapper. According to the intended generation formation, the wrapper technique will iterate (El Aboudi & Benhlima, 2016). The SVM is the model object that is utilized; it is capable of resolving complicated issues like nonlinear data and high data dimensions (Zhou et al., 2019).

The ensemble approach will be used to train the outcomes of the subset of features and data that have attained the designated generation (stop criterion). Accuracy values are generated by the ensemble technique, which trains the data in parallel (Dong et al., 2020). "Stacking" is one approach used in the ensemble method (Jayapermana et al., 2022). Depending on the machine learning method being utilized, the stacking technique can train data. Training will be separated into base and meta-learners by the stacking method (Dou et al., 2020). At the base learner stage, a preset algorithm will be used to train the data. Three machine learning algorithms—a decision tree (Zaman et al., 2020), a k-nearest neighbor (Wang & Liu, 2021), and a light gradient boosting machine (Ke et al., 2017)—are used as the basis learner in this study. Extreme gradient boosting is employed to forecast bankruptcy at the meta-learner stage (Khoirunnisa et al., 2021; Muslim & Dasril, 2021). Despite not using the filtering method, Zelenkov et al. (2017) used the ensemble classifier. AdaBoost, an ensemble learning technique, was used by Barboza et al. (2017) to lower the error rate when choosing FRs attributes. In order to determine the optimal features, Liang et al. (2016) employed SVM with linear kernel parameters after using GA as feature selection to reduce the dimensions of the data. Dasril and the Muslim

Using k-nearest neighbor, decision tree, gradient boosting trees, and random forest as the basic learner, (2021) employs the extreme gradient boosting approach for feature selection and ensemble stacking. A light gradient boosting machine is the meta-learner that is employed. Li et al. (2017) are among the several studies that have been carried out using bankruptcy records. It gathers data on small and medium-sized businesses in Italy and analyzes insolvency using McNemar validation. Liang et al. (2016) frequently use the Taiwanese dataset, Lin et al. (2019) compare the single and ensemble learning-based bankruptcy, and Muslim and Dasril (2021) and Fernández et al. (2018) use the bankruptcy dataset for Poland.

The ensemble technique, which has been used by Muslim and Dasril (2021) with Polish bankruptcy data, Pisula (2020) with Poland bankruptcy data, and Guo et al. (2022) with Taiwanese bankruptcy data, is also used in research pertaining to bankruptcy prediction. With various meta-learner variants and feature selection techniques, each of these studies suggests the ensemble learning approach. As a result, this study is suggested based on related research sources.

Prior studies on bankruptcy prediction did not attempt to balance class data and instead concentrated solely on improving the classification model's accuracy while ignoring the overabundance of variable data dimensions. This research offers a way for handling unbalanced data and a wrapper methodology for choosing the best features, even if the classification model used is derived on multiple references to previous bankruptcy prediction research, including the stacking ensemble method. The goal of improving this approach is to increase the performance of bankruptcy forecast accuracy.

In this study, we use the Taiwanese Bankruptcy dataset from the Taiwan Economic Journal to increase the accuracy of bankruptcy predictions using a genetic algorithm-support vector machine (GA-SVM) and stacking ensemble approach. The manuscript's remaining sections are arranged as follows. The works pertaining to bankruptcy, feature selection models, and classification models are reviewed in Section 2. The GA-SVM approach, stacking ensemble model, and further machine learning models are presented in Section 3. The results are analyzed and discussed in Section 4. Conclusion of Section 5.

2. Literature Review

Studies like those by du Jardin (2016), He et al. (2018), and Kliestik et al. (2020) that use a classification model to predict bankruptcy are based on statistical data supplied by financial firms. Multivariant discriminant analysis, logistic regression, neural networks, support vector machines (SVM), and ensemble methods are examples of classification techniques that can resolve classification issues (Qu et al., 2019). Other neural network-based algorithms, like the convolutional neural network (Hosaka, 2019) and recurrent neural network (Chong et al., 2017), have been created to assess finance and management topics (Qu et al., 2019). To improve prediction accuracy, high indicator dimensions can be reduced through feature selection (Bouktif et al., 2018; Pampouktsi et al., 2023). According to Linero (2018), because the properties of the data are unimportant to the algorithm, huge dimensions can worsen the performance of the resultant accuracy forecast, as in the case of decision trees and Naïve Bayes. According to Liang et al. (2016), feature selection can lower the high dimensions derived from the combination of FRs and CGIs indicators.

For classification issues, the feature selection strategy has been widely optimized (Dessì & Pes, 2015; Tran et al., 2018; Zhang et al., 2018). The impact of filter and wrapper-based feature selection techniques in bankruptcy prediction was investigated by Liang et al. (2016). In experimental datasets, they were unable to identify the best mix of feature selection and classification algorithms. According to the empirical findings, the suggested wrapper approach outperforms conventional feature selection models in terms of prediction accuracy.

The use of factors to assess financial bankruptcy that may come to more than 20 results in a dataset with large dimensions, such as the Taiwanese dataset assembled by Liang et al. (2016). Feature selection approaches including GA (Liang et al., 2016; Lin et al., 2019; Prasetiyo et al., 2021), extreme gradient boosting (Muslim & Dasril, 2021), and SVM (Lin et al., 2019) are commonly used since large dataset sizes might make it difficult to compile bankruptcy prediction assessments.

Although feature selection has helped with efforts to improve prediction accuracy, it is not enough. Machine learning techniques for dataset training included single classifier GA (Ajani et al., 2021; Li et al., 2017; Lin et al., 2019), logistic regression (Li et al., 2017), k-nearest (Noori et al., 2017), SVM (Liang et al., 2016; Lin et al., 2019; Prasetiyo et al., 2021; Qu et al., 2019), and light gradient boosting machine (Zelenkov et al., 2017). A stacking ensemble can be used to increase the accuracy of any classification technique (Muslim & Dasril, 2021).

Using a dataset of Taiwanese companies, Brenes et al. (2022) developed a multilayer perceptron model to forecast bankruptcy with an accuracy of 86.06%. In a related study, 97% accuracy was achieved by combining the XGBoost feature selection algorithm with the stacking ensemble (Muslim & Dasril, 2021). Nevertheless, based on earlier studies, approaches still need to be modified to achieve the highest level of accuracy. Therefore, the researcher discovered a method to identify a method to predict bankruptcy with high accuracy from prior studies based on the relevant job described above.

3. Quantum Research Methodology

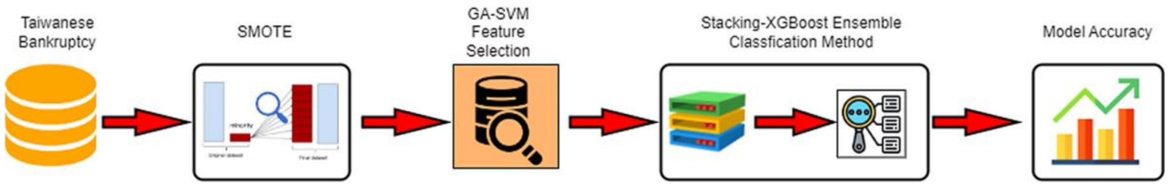
3.1 Research design

The purpose of the research design was to provide a thorough explanation of the research flow. Three stages are often described by the workflow: preprocessing, feature selection, and data training. A validation test employing a confusion matrix and k-fold cross-validation concluded the investigation. Figure 1 illustrates how the suggested approach is visualized.

3.1.1. Description of the data

For the investigation, a random sample technique was used. Twenty-two first-year in-service postgraduate science instructors from one of Bhutan's education colleges made up the study's sample. Nine female teachers (40.9%) and thirteen male teachers (59%) took part in the survey. Five physics instructors (22.7%), ten chemistry teachers (45.5%), and seven biology teachers (31.8%) made up the sample. Only 22 of the 39 in-service postgraduate science professors took part in the study because it was entirely voluntary. A total of 56% of respondents responded.

Figure 1: The proposed method of bankruptcy prediction with GA-SVM and stacking



3.1.2. The step of preprocessing

According to Li et al. (2017), the preprocessing stage attempts to prepare the dataset for algorithm training. The dataset identified anomalies, noise, and dataset balance. Low precision and an imbalanced dataset can have an impact on the training process (Fernández et al., 2018). The synthetic minority over-sampling technique is the preprocessing step employed in this investigation. In order to create balanced data based on the distribution of the closest k value, the synthetic minority over-sampling technique can be applied to noisy datasets (Prasetiyo et al., 2021).

3.1.3. Stage of feature selection

Depending on the features employed, the feature selection stage seeks to minimize the dataset's dimensions. GA and SVM are combined in feature selection. The two algorithms are combined in a wrapper (El Aboudi & Benhlima, 2016). Based on the generated generating iteration, the GA parameters apply the stopping criterion. According to Liang et al. (2016), the population size is 70, the crossover rate is 0.7, the mutation rate is 0.01 and the generation utilized is 20. Figure 2 provides an example of the GA. Every iteration uses SVM as an evaluator. Following selection, the accuracy of the features is evaluated using an SVM evaluator. The following is the SVM model in pseudocode (Ajani et al., 2021).

Figure 2: Illustration of the feature selection process of GA-SVM

The algorithm

1. SVM Model

Require: X and y loaded with labeled data, α 0
1: C <= some value (use 100, for example)
2: Kernel $\rightarrow$ formula (RBF, for example)
3: repeat
4: for all $\{x_i; y_i\}, \{x_j; y_j\}$ do
5: Optimize α_i and α_j
6: end for
7: until no changes in α or other resource constraint criteria are met
Ensure: Retain only the support vector $(\alpha_i > 0)$

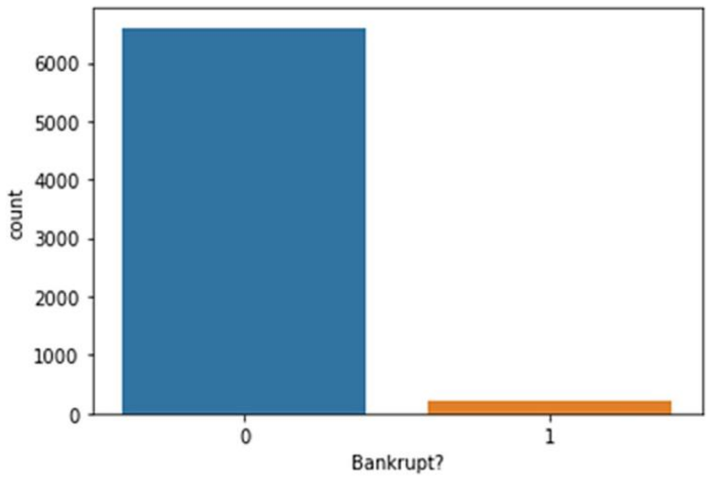
3.1.4. Stacking group education

"Stacking" is one approach used in the ensemble method (Jayapermana et al., 2022). Depending on the machine learning method being utilized, the stacking technique can train data. To create a more realistic model, the ensemble technique integrates several models. According to Abdar et al. (2020), the ensemble approach can greatly reduce misclassification and increase the effectiveness of a single classification model. The ensemble technique combines different sets of models to create a more accurate model. The efficacy of a single classifier can be increased and misclassification can be significantly decreased with the ensemble technique. The algorithm proposes two key ideas: the basic model and the meta-learner. The basic idea is to train the meta-learner using the prediction result of the base model as a new feature matrix, and then utilize the meta-learner's output as the final prediction result after combining the techniques. The output of one layer serves as the input for the layer behind it in the multilayer stacking ensemble learning model. However, the model gets more complicated and the training rate decreases as the number of layers increases. Three machine learning algorithms—a decision tree (Zaman et al., 2020), a k-nearest neighbor (Wang & Liu, 2021), and a light gradient boosting machine (Ke et al., 2017)—are used as the basis learner in this study. Extreme gradient boosting is employed to forecast bankruptcy at the meta-learner stage (Khoirunnisa et al., 2021; Muslim & Dasril, 2021).

4. Result and Discussion

According to the study's dataset, the ratio of label 1 to label 0 for the goal classification label "Bankrupt?" is 96.77%:3.23%. According to the original dataset, 220 instances have the target label 1 and up to 6599 instances have the target label 0. Figure 3 shows the depiction of the target label comparison.

Figure 3: Original dataset label target of Taiwanese Bankruptcy dataset



Since it would impact the data training accuracy, the unbalanced data target in Figure 3 needs to be improved. resampling data to the closest k value in the dataset's distribution in order to correct unbalanced data. The majority label is described as "Bankrupt?" 0, whereas the minority label is described as "Bankrupt?" 1. A synthetic over-sampling technique will be used to resample the dataset with a minority label (Prasetiyo et al., 2021). The resampling procedure begins with 50% of the dataset from the majority label and continues until 100% of the dataset is resampled from the majority label. Figure 4 shows a visualization of the synthetic minority over-sampling method used.

Figure 4: The use of the synthetic minority oversampling technique method. Figure (a) is the original data, (b) resampling 50% of the majority label, and (c) resampling 100% of the majority label

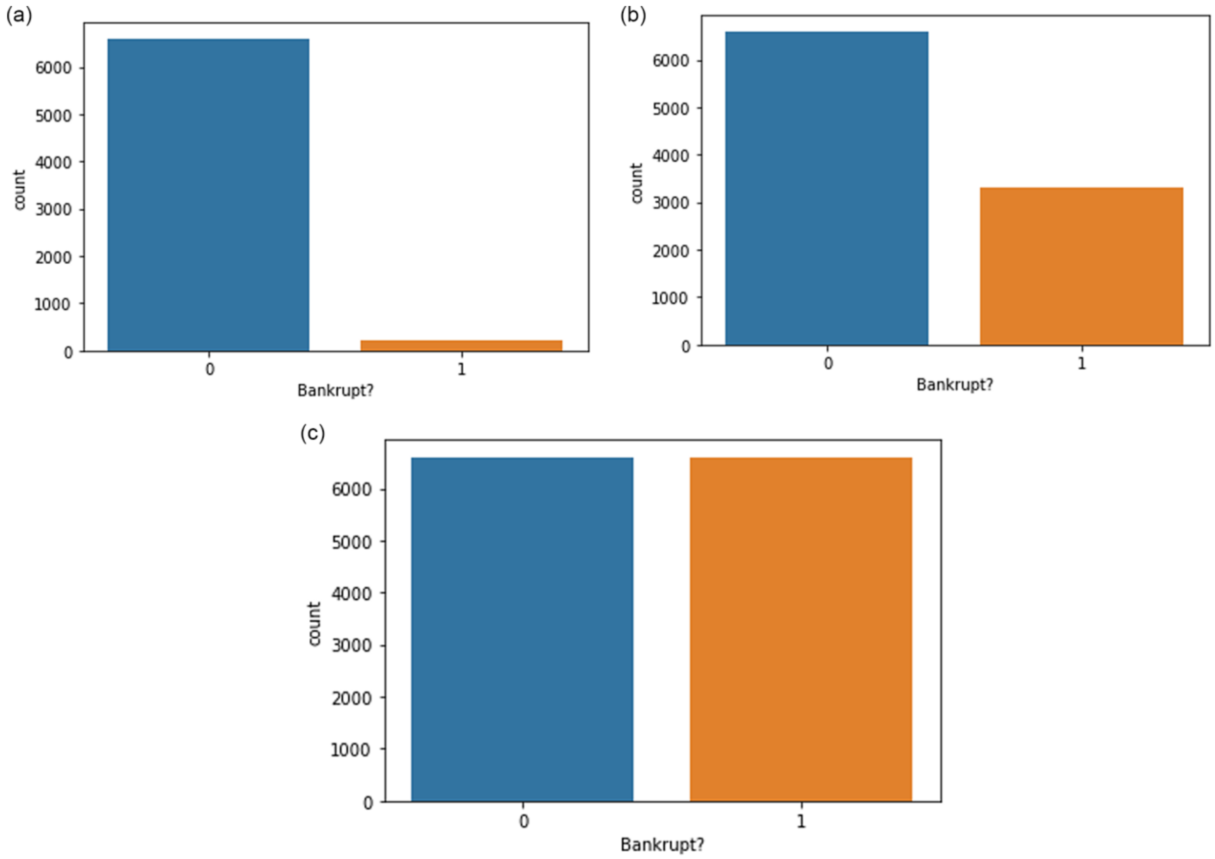
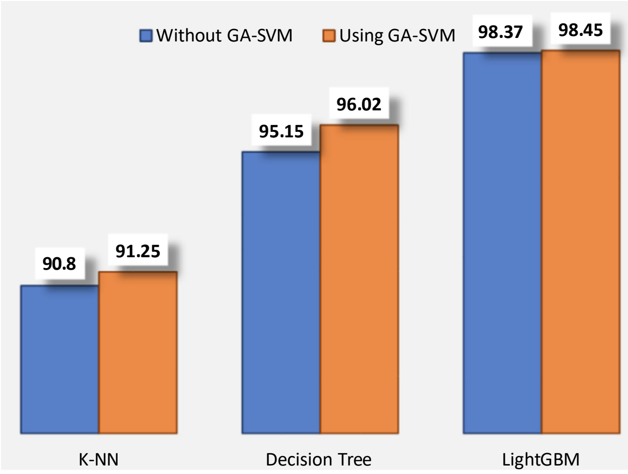


Table 1 shows how many resampling results were produced using the given %. Minority labels are resampled by 100% of the majority labels in the new dataset used for data training. As a result, balanced data was used for data training. GA-SVM reduces used features by processing balanced data. Gokulnath and Shantharajah's (2019) study extracted seven features from heart disease data and produced an accuracy of 83.70%, whereas Noori et al. (2017) used fNIRS signal data to derive three features of accuracy at 91.00%. Taiwanese bankruptcy data was collected in 43 cases using the GA-SVM approach for feature selection in this study, with an accuracy of 92.05%, as indicated in Table 2.

Out of 92 native features, GA-SVM chose 43 as the greatest features. Various accuracy levels are produced as a result of the training data procedure. The techniques are light gradient boosting machine, k-nearest neighbor, and decision tree. The data training methodology employs stacking in addition to a single classifier. Table 3 shows the outcome of the comparison between the stacking approach utilizing meta-learner extreme gradient boosting and training data testing using a single classifier that has been identified (Hou et al., 2021). Using a decision tree for data training without feature selection yields 95.15%, a k-nearest neighbor yields 90.8%, and a light gradient boosting machine yields 98.37%. The light gradient boosting machine yields 98.45%, the k-nearest neighbor yields 91.25%, and the GA-SVM feature selection utilizing a decision tree yields 96.02%. to the findings of Tao et al.'s (2018) study.

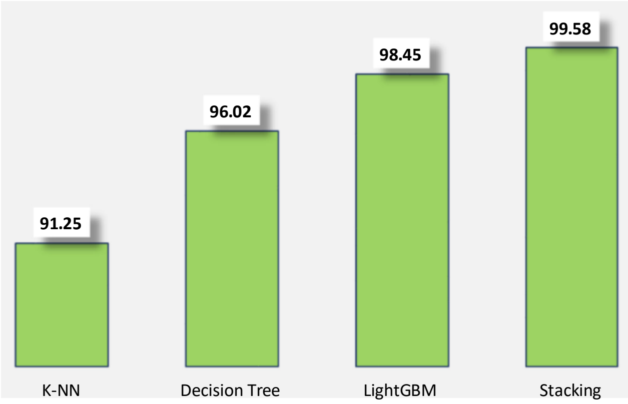
The accuracy of each classifier's feature selection results utilizing GA-SVM has improved. as illustrated in Figure 5. The decision tree, k-nearest neighbor, and light gradient boosting machine are used as foundation learners in the ensemble stacking technique. The base learner yielded a new dataset that categorizes using meta-learning. Extreme gradient boosting classification is the meta-learner employed in the ensemble stacking technique. The comparison between the ensemble classifier (extreme gradient boosting with stacked decision tree, k-nearest neighbor, and light gradient boosting machine) and single classifier (decision tree, k-nearest neighbor, and light gradient boosting machine) displays the findings. Table 4.

Figure 5: Comparison of single classifier without GA-SVM and with GA-SVM



A decision tree, k-nearest neighbor, and light gradient boosting machine are stacked with an ensemble classifier as a base learner and an extreme gradient boosting as a meta-learner to reach a 99.58% accuracy. Figure 6 shows the comparison visualization between single classifier and stacking.

Figure 6: Comparison of single classifier and stacking



Islamic and Dasril's (2021) study, which used Polish bankruptcy data and ensemble stacking, produced findings of 97%, whereas Pisula's (2020) study, which used Poland bankruptcy data ensemble stacking, produced 98.1% accuracy. Furthermore, a related experiment by Guo et al. (2022) tested the bagging ensemble model with the Taiwanese Bankruptcy dataset and achieved an accuracy of 86.63%. Therefore, using Taiwanese bankruptcy data, it can be demonstrated that the suggested stacking ensemble learning approach is better.

5. Conclusion

Research is done to analyze machine learning predictions for bankruptcy. The study employed extreme gradient boosting as a meta-learner, a single classifier as a base learner, and the GA-SVM machine learning algorithm for feature selection. The single classifier serves as a light gradient boosting machine, k-nearest neighbor, and decision tree. Following the selection of characteristics using GA-SVM through a single classifier, the dataset training results show that the decision tree generates 96.02%, the k-nearest neighbor generates 91.25%, and the light gradient boosting machine generates 98.45%. 99.58% was the outcome of dataset training with the ensemble stacking approach. It has been demonstrated that employing a stacking ensemble with meta-learner extreme gradient boosting can improve the bankruptcy classification accuracy value.

References

Abdar, M., Zomorodi-Moghadam, M., Zhou, X., Gururajan, R., Tao, X., Barua, P. D., & Gururajan, R. (2020). A new nested ensemble technique for automated diagnosis of breast cancer. *Pattern Recognition Letters*, *132*, 123–131. https://doi.org/10.1016/ j.patrec.2018.11.004

Abdillah, L. A. (2020). FinTech E-commerce payment application user experience analysis during COVID-19 pandemic. *arXiv Preprint: 2012.07750*. https://doi.org/10.48550/arXiv.2012.07750

Ajani, T. S., Imoize, A. L., & Atayero, A. A. (2021). An overview of machine learning within embedded and mobile devices – Optimizations and applications. *Sensors*, *21*(13), 4412. https://doi.org/10.3390/s21134412

Almamy, J., Aston, J., & Ngwa, L. N. (2016). An evaluation of Altman's Z-score using cash flow ratio to predict corporate failure amid the recent financial crisis: Evidence from the UK. *Journal of Corporate Finance*, *36*, 278–285. https://doi. org/10.1016/j.jcorpfin.2015.12.009

Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, *83*, 405–417. https://doi.org/10.1016/j.eswa.2017.04.006

Bateni, L., & Asghari, F. (2020). Bankruptcy prediction using logit and genetic algorithm models: A comparative analysis. *Computational Economics*, *55*, 335–348. https://doi.org/10. 1007/s10614-016-9590-3

Blazy, R., & Stef, N. (2020). Bankruptcy procedures in the post-transition economies. *European Journal of Law and Economics*, *50*, 7–64. https://doi.org/10.1007/s10657-019-09634-5

Bouktif, S., Fiaz, A., Ouni, A., & Serhani, M. A. (2018). Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies*, *11*(7), 1636. https://doi.org/10.3390/en11071636

Brenes, R. F., Johannssen, A., & Chukhrova, N. (2022). An intelligent bankruptcy prediction model using a multi- layer perceptron. *Intelligent Systems with Applications*, *16*, 200136. https://doi.org/10.1016/j.iswa.2022.200136

Chong, E., Han, C., & Park, F. C. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems with Applications*, *83*, 187–205. https://doi.org/10.1016/j.eswa. 2017.04.030

Dessì, N., & Pes, B. (2015). Similarity of feature selection methods: An empirical study across data intensive classification tasks. *Expert Systems with Applications*, *42*(10), 4632–4642. https://doi.org/10.1016/j.eswa.2015.01.069

Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. *Frontiers of Computer Science*, *14*, 241–258. https://doi.org/10.1007/s11704-019-8208-z

Dou, J., Yunus, A. P., Bui, D. T., Merghadi, A., Sahana, M., Zhu, Z., … , & Pham, B. T. (2020). Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan. *Landslides*, *17*, 641–658. https://doi.org/10.1007/s10346-019-01286-5

du Jardin, P. (2016). A two-stage classification technique for bankruptcy prediction. *European Journal of Operational Research*, *254*(1), 236–252. https://doi.org/10.1016/j.ejor.2016.03.008

El Aboudi, N., & Benhlima, L. (2016). Review on wrapper feature selection approaches. In *2016 International Conference on Engineering & MIS*, 1–5. https://doi.org/10.1109/ICEMIS. 2016.7745366

Fernández, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary. *Journal of Artificial Intelligence Research*, *61*, 863–905. https://doi.org/ 10.1613/jair.1.11192

Foerster, S., Tsagarelis, J., & Wang, G. (2017). Are cash flows better stock return predictors than profits? *Financial Analysts Journal*, *73*(1), 73–99. https://doi.org/10.2469/faj.v73.n1.2

Gokulnath, C. B., & Shantharajah, S. P. (2019). An optimized feature selection based on genetic approach and support vector machine for heart disease. *Cluster Computing*, *22*, 14777–14787. https://doi.org/10.1007/s10586-018-2416-4

Guo, Y., Mei, J., Pan, Z., Liu, H., & Li, W. (2022). Adaptively promoting diversity in a novel ensemble method for imbalanced credit-risk evaluation. *Mathematics*, *10*(11), 1790. https://doi.org/10.3390/ math10111790

He, H., Zhang, W., & Zhang, S. (2018). A novel ensemble method for credit scoring: Adaption of different imbalance ratios. *Expert Systems with Applications*, *98*, 105–117. https://doi. org/10.1016/j.eswa.2018.01.012

Hickel, J., Dorninger, C., Wieland, H., & Suwandi, I. (2022). Imperialist appropriation in the world economy: Drain from the global south through unequal exchange, 1990–2015. *Global Environmental Change*, *73*, 102467. https://doi.org/ 10.1016/j.gloenvcha.2022.102467

Hosaka, T. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks. *Expert Systems with Applications*, *117*, 287–299. https://doi.org/10.1016/ j.eswa.2018.09.039

Hou, H., Chen, X., Li, M., Zhu, L., Huang, Y., & Yu, J. (2021).

Prediction of user outage under typhoon disaster based on multi-algorithm Stacking integration. *International Journal of Electrical Power & Energy Systems*, *131*, 107123. https:// doi.org/10.1016/j.ijepes.2021.107123

Jayapermana, R., Aradea, A., & Kurniati, N. I. (2022). Implementation of stacking ensemble classifier for multi-class classification of COVID-19 vaccines topics on Twitter. *Scientific Journal of Informatics*, *9*(1), 8–15.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., . . . , & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In *31st Conference on Neural Information Processing Systems*.

Khoirunnisa, O., Pertiwi, D. A. A., Dianti, E. N., Fattah, A. M. M., & Budiman, K. (2021). Improvement business process model and notation on the drink distribution industries using six core element. *Journal of Soft Computing Exploration*, *2*(2), 99–106. https://doi.org/10.52465/joscex.v2i2.46

Kliestik, T., Valaskova, K., Lazaroiu, G., Kovacova, M., & Vrbka, J. (2020). Remaining financially healthy and competitive: The role of financial predictors. *Journal of Competitiveness*, *12*(1), 74–92. https://doi.org/10.7441/joc.2020.01.05

Kozlovskyi, S., Poliakov, B., Lavrov, R., & Ivanyuta, N. (2019). Management and comprehensive assessment of the probability of bankruptcy of Ukrainian enterprises based on the methods of fuzzy sets theory. *Problems and Perspectives in Management*, *17*(3), 370–381. http://dx.doi.org/10.21511/ppm.17(3).2019.30

Kücher, A., Mayr, S., Mitter, C., Duller, C., & Feldbauer-Durstmüller, B. (2020). Firm age dynamics and causes of corporate bankruptcy: Age dependent explanations for business failure. *Review of Managerial Science*, *14*, 633–661. https://doi.org/10.1007/ s11846-018-0303-2

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2017). Feature selection: A data perspective. *ACM Computing Surveys*, *50*(6), 94. https://doi.org/10.1145/3136625

Liang, D., Lu, C. C., Tsai, C. F., & Shih, G. A. (2016). Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. *European Journal of Operational Research*, *252*(2), 561–572. https://doi.org/10. 1016/j.ejor.2016.01.012

Liang, D., Tsai, C. F., Lu, H. Y., & Chang, L. S. (2020). Combining corporate governance indicators with stacking ensembles for financial distress prediction. *Journal of Business Research*, *120*, 137–146. https://doi.org/10.1016/j.jbusres.2020.07.052

Lin, W. C., Lu, Y. H., & Tsai, C. F. (2019). Feature selection in single and ensemble learning-based bankruptcy prediction models. *Expert Systems*, *36*(1), e12335. https://doi.org/10.1111/exsy.12335

Linero, A. R. (2018). Bayesian regression trees for high-dimensional prediction and variable selection. *Journal of the American Statistical Association*, *113*(522), 626–636. https://doi.org/ 10.1080/01621459.2016.1264957

Muslim, M. A., & Dasril, Y. (2021). Company bankruptcy prediction framework based on the most influential features using XGBoost and stacking ensemble learning. *International Journal of Electrical and Computer Engineering*, *11*(6), 5549–5557. https://doi.org/10.11591/ijece.v11i6.pp5549-5557

Jumanto, J., Muslim, M. A., Dasril, Y., & Mustaqim, T. (2023). Accuracy of Malaysia public response to economic factors during the Covid-19 pandemic using vader and random. *Journal of Information System Exploration and Research*, *1*(1), 49–70. https://doi.org/10.52465/joiser.v1i1.104

Noori, F. M., Naseer, N., Qureshi, N. K., Nazeer, H., & Khan, R. A. (2017). Optimal feature selection from fNIRS signals using genetic algorithms for BCI. *Neuroscience Letters*, *647*, 61–66. https://doi.org/10.1016/j.neulet.2017.03.013

Pampouktsi, P., Avdimiotis, S., Maragoudakis, M., Avlonitis, M., Samantha, N., Hoogar, P., . . . , & Rono, W. (2023). Techniques of applied machine learning being utilized for the purpose of selecting and placing human resources within the public sector. *Journal of Information System Exploration and Research*, *1*(1), 1–16. https://doi.org/10.52465/joiser.v1i1.91

Pisula, T. (2020). An ensemble classifier-based scoring model for predicting bankruptcy of polish companies in the Podkarpackie Voivodeship. *Journal of Risk and Financial Management*, *13*(2), 37. https://doi.org/10.3390/jrfm13020037

Prasetiyo, B., Muslim, M. A., & Baroroh, N. (2021). Evaluation performance recall and F2 score of credit card fraud detection unbalanced dataset using SMOTE oversampling technique. *Journal of Physics: Conference Series*, *1918*, 42002. https:// doi.org/10.1088/1742-6596/1918/4/042002

Qu, Y., Quan, P., Lei, M., & Shi, Y. (2019). Review of bankruptcy prediction using machine learning and deep learning techniques. *Procedia Computer Science*, *162*, 895–899. https://doi.org/10.1016/j.procs.2019.12.065

Tao, P., Sun, Z., & Sun, Z. (2018). An improved intrusion detection algorithm based on GA and SVM. *IEEE Access*, *6*, 13624–13631. https://doi.org/10.1109/ACCESS.2018.2810198

Zhou, T., Lu, H., Wang, W., & Yong, Y. (2019). GA-SVM based feature selection and parameter optimization in hospitalization expense modeling. *Applied Soft Computing*, *75*, 323–332. https://doi.org/10.1016/j.asoc.2018.11.001

Tran, C. T., Zhang, M., Andreae, P., Xue, B., & Bui, L. T. (2018). Improving performance of classification on incomplete data using feature selection and clustering. *Applied Soft Computing*, *73*, 848–861. https://doi.org/10.1016/j.asoc.2018.09.026

Umar, M., Ji, X., Mirza, N., & Naqvi, B. (2021). Carbon neutrality, bank lending, and credit risk: Evidence from the Eurozone. *Journal of Environmental Management*, *296*, 113156. https://doi.org/10.1016/j.jenvman.2021.113156

Wang, H., & Liu, X. (2021). Undersampling bankruptcy prediction: Taiwan bankruptcy data. *PLOS ONE*, *16*(7), e0254030. https:// doi.org/10.1371/journal.pone.0254030

Zaman, M., Kaul, S., & Ahmed, M. (2020). Analytical comparison between the information gain and Gini index using historical geographical data. *International Journal of Advanced Computer Science and Applications*, *11*.

Zelenkov, Y., Fedorova, E., & Chekrizov, D. (2017). Two-step classification method based on genetic algorithm for bankruptcy forecasting. *Expert Systems with Applications*, *88*, 393–401. https://doi.org/10.1016/j.eswa.2017.07.025

Zhang, L., Mistry, K., Lim, C. P., & Neoh, S. C. (2018). Feature selection using firefly optimization for classification and regression models. *Decision Support Systems*, *106*, 64–85. https://doi.org/10.1016/j.dss.2017.12.001